DOCUMENT RESUME

ED 473 068 HE 035 659

AUTHOR	Cohoon, J. McGrath
TITLE	Data for Multi-Institution Research. AIR 2002 Forum Paper.
PUB DATE	2002-06-00
NOTE	31p.; Paper presented at the Annual Meeting of the Association for Institutional Research (42nd, Toronto, Ontario, Canada, June 2-5, 2002).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC02 Plus Postage.
DESCRIPTORS	Case Studies; *Colleges; *Data Collection; Higher Education; *Institutional Characteristics; Sex Differences; *Student Attrition
IDENTIFIERS	*Multiple Site Studies

ABSTRACT

Although multi-institutional data are more available than ever, researchers still face daunting hurdles when collecting data from multiple institutions. These hurdles are particularly high for stand-alone projects. This paper makes the case that, despite the difficulties, results from multi-institution research benefit both participating colleges or universities and higher education as a whole. In support of this argument, a detailed case study of the data collection for one multi-institutional research project is reported. This case study describes an investigation into the departmental factors associated with disproportionate loss of women from undergraduate computing majors. The case study identified data source, illustrated challenges to collecting data from numerous institutions, and noted potential rewards for both higher education and participating institutions. (Contains 2 figures and 16 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 473 068

Data for Multi-Institution Research

J. McGrath Cohoon
University of Virginia

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

D. Vura

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

Abstract

Although multi-institution data is more available than ever, researchers still face daunting hurdles when collecting data from multiple institutions. These hurdles are particularly high for stand-alone projects. This paper argues that despite the difficulties, results from multi-institution research benefit both participating colleges or universities and higher education as a whole.

In support of this argument, a detailed case study of the data collection for one multi-institutional research project is reported. This case study describes an investigation into the departmental factors associated with disproportionate loss of women from undergraduate computing majors. The case study identifies data sources, illustrates challenges to collecting data from numerous institutions, and notes potential rewards for both higher education and participating institutions.

Data for Multi-Institution Research

Multi-institution research benefits both participating institutions and higher education as a whole. When colleges and universities join together to investigate common concerns, institutions gain information that enhances program evaluation, planning, and decision-making, particularly regarding their position relative to other institutions (Trainer, 1996). The money spent on participating in multi-institution research saves participating institutions from the high cost of uninformed decisions (Hackett, 1996). Higher education as a whole benefits from multi-institution research because it gains information that both satisfies demands for accountability and increases understanding of important issues. Examples of the issues addressed by multi-institution research include racial equity (Pavel and Reiser, 1991), institutional quality (Vinsonhaler and Vinsonhaler, 1991), and financial concerns (Creswell, Chronister, Brown, 1991). These examples illustrate how the broad interests of society and higher education advance when generalizable findings and comparative measures are produced through research involving many institutions.

Technological progress has improved the processes for collecting original data from large populations, (for example, see Dillman, 2001). Technology has also facilitated implementation of student tracking systems (Borden, 1995) and the resulting improvements in the collection of secondary data from multiple institutions. Official data repositories such as state higher education councils, WebCASPAR, data-sharing consortiums, and offices of institutional research are outstanding sources of data from multiple institutions. In particular, these large student unit

databases are often well established, contain a wealth of information, and cover broad geographical areas. Their worth and increasing prevalence in the United States are documented (Russell and Chisholm, 1995), and can be observed by visiting websites such as AIR's "Internet Resources for Institutional Research" at http://airweb.org/links/internetreports.html. By 1999, forty-two states had postsecondary education data systems, most of which were unit record databases (Russell, 1999.) Their usability improves over time as easily demonstrated by comparing WebCASPAR's functionality today with that described by Firnberg in 1991. However, when a project's needs do not fit with statistics that have already been produced, the persisting limitations of these data sources become apparent.

Regardless of the benefits and recent advances in methodology, there are still hurdles that must be overcome when gathering data for multi-institution research. Difficulties occur when data are unavailable from official sources, a situation that occurs for a variety of reasons. National and statewide databases may not contain necessary information because the data were not collected. Even if the desired information does exit, it may not be available because the storage format can make data difficult to extract. Likewise, proprietors of institutional databases are not always able or willing to accommodate external requests. Sapp (1996) suggested possible reasons for reluctance in relation to data-sharing consortia: the cost associated with personnel and time; concerns over whether institutional comparisons are valid or potentially unfavorable; and concerns about sensitive data.

The following case study illustrates the current state of multi-institutional research. It specifies data sources, methods, hurdles, and the potential rewards to higher education and

participating institutions. This detailed example is presented to encourage continued development of data resources and increased access to data, and to promote broad support for valuable multi-institution research.

Case study of a multi-institution research project

"Departmental Factors in Gendered Attrition from Undergraduate IT Majors," is a study of computer science program retention by gender that began in the fall of 2000. This three-year project, funded by the National Science Foundation, was built upon a statewide study in Virginia. Expanding on the pilot, the nationwide study examines a timely and important issue that is relevant to both our nation's economic well-being and gender equity. The study is an attempt to determine why many, but not all, undergraduate computing programs lose women to other majors at higher rates than they lose men. In a similar fashion to Young and Reglinger's (2000) focus on student flows through academic majors, this study tracks the outflow of students from computing majors. The results will identify which departmental factors influence the size of the gender gap in attrition from this discipline.

The focus of the study is on department-level influences, as opposed to gender or socialization influences on students. The unit of analysis is the department, so we take a novel approach to calculating our dependent variable – gendered attrition rates. Gendered attrition rates are the difference between a department's male and female average annual rates of migration out during the six study years. By comparing attrition rates for men and women within the same department, the gendered attrition rate measures the difference in outcomes for two groups in the

same environment. Thus, this measure is a useful tool for investigating gendered outcomes across institutions because it controls for departments' overall attrition rates.

Data from surveys and interviews are aggregated by department to measure most of the independent variables. These variables include numerous aspects of departmental resources and faculty practices and attitudes. Aggregate data from official records of enrollment and graduation measure the dependent variable and the gender composition independent variable.

Two hundred ten study departments were selected because they rank among the most prestigious computer science programs and/or they recently graduated the largest numbers of computer science baccalaureates in the contiguous United States. We identified these departments based on 1996 and 1997 degrees-granted data from the National Science Foundation's WebCASPAR website (http://caspar.nsf.gov.) As a group, the study departments produce approximately three fifths of the computer science bachelors degrees in the nation.

Efforts to obtain the data for the dependent variable have been in progress since the project began. Acquiring these data has been the most challenging aspect of the project. To overcome low initial response, we successively contacted several potential data sources for each study institution. These sources included government data repositories, sponsors within the study institutions, and institutional researchers. None of these approaches has yet been highly successful, although each has yielded some results. Efforts will continue until analyses are well underway, so that the maximum number of institutions can be included.

The next two sections describe how the interview and survey data were obtained from individual respondents at various institutions. These tasks were undertaken during the spring

semesters of 2001 and 2002. The third section describes the various approaches to obtaining data from official sources.

INTERVIEWS AT 18 INSTITUTIONS

The goal of the interview phase of the project was to collect qualitative data from multiple institutions. These data would ground the study in the realities of undergraduate computer science education and describe participants' experiences in their own words. The interview phase was highly successful, although there were significant costs and logistic considerations. The results were valuable both as research products and as enhancements to the quality of the overall project.

Interviewer teams traveled to four urban locations across the United States and three non-urban locations in Virginia. The urban sites were selected for variety in geographic location and for availability of several study institutions in each locale. The selected urban locations were New York City, Chicago, San Diego/Los Angeles, and Atlanta. At each of these locations, specific computer science departments were selected for variety in gender balance and institutional type. The non-urban sites were a convenience sample.

Once sites were selected, department chairpersons were contacted by letter and by telephone to request permission for a visit. Cooperation was high – 18 of the 22 departments (82%) contacted agreed to be visited. However, the number of people who participated at each site did not always meet expectations. We maximized faculty and student participation by scheduling multiple researchers in a department at one time and spread over an entire week.

Student focus group participants were offered pizza and soda, and in a few cases, departments provided incentives such as t-shirts.

Four interviewers (the Principal Investigator and three trained, experienced graduate students) traveled to each site. During each trip, interviewers visited three or four institutions to interview all willing participants. In almost every case, the undergraduates were interviewed in sex-segregated focus groups. Faculty and administrator interviews were one-on-one and semi-structured. As a result of this process, the interviewers spoke with 325 members of undergraduate computer science programs: 23 faculty who were chairpersons or administrators, 120 teaching faculty, and 182 undergraduates in 31 groups. Participants were all recruited and scheduled by the chairperson or the chairperson's designated representative.

Interviewers asked chairpersons about program and faculty characteristics, teaching emphasis, resources, and their observations regarding the relationship between students' sex and retention in their computing program. Faculty were prompted to speak about their teaching and mentoring, the environment in the department, and their observations of the relationship between students' sex and retention. Students were asked to discuss what drew them to computing and their particular program, positive and negative experiences, and how they cope with negative experiences.

All but two participants agreed to be audio-taped; untaped interviews were recoded with interviewer notes. Audio tapes were copied and transcribed, and the transcriptions were reviewed for accuracy. Analysis of the transcripts and notes is continuing with particular attention to differences between departments. The analyses have already produced some interesting

preliminary results on issues such as variation in pedagogical practices, the gendered nature of student satisfaction with teaching, attitudes and behaviors with respect to academic dishonesty, and the appeal of computer science.

The qualitative component of the project contributed significantly to the overall project. Visiting a variety of institutions made it possible to compare departments, leading to observations of some surprising commonalities. For example, students frequently raised the issue of cheating. The intensity of their feelings and the degree to which they found cheating to be discouraging were issues not noted in the literature on student retention in computing majors. The relationship between academic dishonesty and student outflow from computing majors bears further examination. Another contribution of the broadly based qualitative data was to inform the construction of the survey that was subsequently sent to all study institutions. Both questions and response categories were improved as a result of findings from the interviews and focus groups.

The challenges involved with this phase of the project included the expense of travel and interviewer pay, coordination of travel and site arrangements, and quality control of the interviews and focus groups, including transcription and transcript coding. Respectively, these challenges were overcome with grant funds, patient and persistent communication and negotiation, and close review of all work.

The rewards for participating institutions were confidential summary reports describing the various characteristics and practices of each interview site. These summary reports contained descriptive information on student reasons for choosing a computing major and their particular program; positive and negative experiences students had in the program; how students cope;

faculty pedagogical and mentoring practices; support programs for students; faculty estimations of student quality and opinions of student qualities necessary for success in the program; faculty evaluations of collegiality, the gender climate in the department, and institutional support the program received; and faculty reasons for considering leaving their position. Comparative reports will be posted on the project website so that departments can evaluate their own interview results relative to other anonymous programs. The benefit to higher education was detailed descriptive knowledge about a variety of conditions and practices in academic departments. This information deepens understanding of student and faculty concerns and how to cope with these concerns.

SURVEYS AT 210 INSTITUTIONS

The goal of the survey phase of the project was to measure characteristics and practices of all study departments. We collected quantitative data from chairpersons and faculty via web, mail, and telephone surveys. The web survey was the first response mode offered; mail was second; and telephone was third. Overall, collection of survey data was highly successful.

To create a sampling frame, contact information for faculty and chairpersons was obtained from the websites of study departments. The information was verified by administrative assistants via telephone conversations or faxed responses. This collection and verification was a very labor-intensive process.

The chairperson of every study department was selected to participate in the survey. Chairperson survey questions addressed features of their undergraduate computing programs,

including emphasis on teaching, resources, concerns, and personal demographic information. In addition to chairpersons, a stratified random sample of up to 25 faculty was selected for each study department with women over-sampled. Faculty questions addressed professional activities and departmental life, focusing on pedagogy, mentoring, evaluations of students and the department, and personal demographic information.

Each selected department head and faculty member was contacted at least ten times before non-response was accepted as a refusal to participate. This effort resulted in an overall response rate of 67% from an eligible sample of 2,526 faculty and 71% for 209 chairpersons. Of the 1683 faculty respondents, 1,514 (90%) responded online, 137 (8%) responded by mail, and 32 (2%) responded by telephone. Of the 152 responding department heads, 138 (91%) responded online, 12 (8%) responded by mail, and 2 (1%) responded by telephone.

Although the survey data are still being prepared for analysis, they have already demonstrated the worth of this multi-institution research. Interesting and thought-provoking results are emerging. For example, the preliminary results from responses via the web show (in Figure 1) that 41% of the 1,502 computer science faculty who responded to this question had seriously considered leaving their position in recent years. The most common motivation for their deliberation was dissatisfaction with the institutional support their department received – 42% of those who considered leaving rated this dissatisfaction as a strong or very strong motivation. Other powerful motivations were dissatisfaction with departmental leadership (37%), the desire for career change or advancement (35%), money (29%), and the desire for better students (27%). Except for retirement, teaching was the weakest stimulus for thoughts of leaving

– only 6% of those who considered leaving were strongly motivated by the desire to teach more; 16% were strongly motivated by the desire to teach less. According to chairpersons, the actual turnover in study departments averaged about 3% annually, and 17% overall during the study period. However, given the opportunities available to computer scientists and the shortage of faculty in this discipline, institutions might regard this potential for turnover as a serious cause for concern.

The challenges involved with the survey phase of this project were the cost of running a large survey, persuading busy faculty and chairpersons to respond, and coordinating efforts with a survey research center hired to implement the mail and telephone components of the project. These challenges were overcome by leading with the web version of the survey, which required no printing, mailing, data entry, or long-distance calling costs; use of the token incentive; polite, but persistent contacts; and close interactions with responsible staff including periodic meetings and frequent progress reports.

The benefits for participating departments were summary data measuring common discipline concerns and teaching and mentoring practices in the average CS department. The benefit to higher education was generalized findings that can inform discussions on subjects such as pedagogical practices, faculty retention, and gender issues in an economically important discipline.

Official Data

The goal of obtaining official enrollment and disposition data was to measure the composition of the student body in each study program, and to calculate the rates at which men and women migrated out of the major during the study period. The official data made it possible to statistically relate measures of departmental conditions and practices with departmental gendered attrition outcomes. Without these official data, the project would produce many interesting and useful findings, but it would not identify which departmental factors significantly influence the size of the gender gap in attrition from computing majors. Unfortunately, obtaining the official data posed the greatest challenge of any aspect of the project.

In order to calculate the rates at which male and female students leave computing majors, we requested that study institutions provide the following data aggregated from student unit records.

> For each year from 1994 through the most recent year available, provide a fall headcount of computer science majors (CIP code 1100) broken down by sex, by level,[1] by CIP sub code, and by full or part-time status. Show average Math SAT score and average major GPA for each group. Track each group to the fall of subsequent years and identify their dispositions. These dispositions are: number still enrolled with the same major, number still enrolled but with a different

[1] Including all levels of students each year overcomes problems with small numbers that can be encountered when tracking out a single cohort of new students. It is comparable to demographers' methods for calculating population death rates.

major, number graduated in the same major, number graduated in a different major, and number no longer enrolled.

These data can be used to calculate measures of programs' gender composition at different points in time, male and female annual rates of migration from the major, and annual and study period average gendered attrition rates (the gap in male and female migration rates).

Government sources

National databases that contain institutional data on enrollment and graduation numbers were the first choice as a source for the information needed to calculate gendered attrition rates. We began with Web CASPAR, the National Science Foundation's online source of statistics on academic science and engineering in the United States. This site provides access to data from a variety of sources including the Integrated Postsecondary Education Data System (IPEDS).

However, after careful investigation, it became clear that national databases do not contain the data needed for quantifying undergraduates' migration out of computing programs. The IPEDS data break down students by major only at graduation; enrollments are not grouped by major, making it impossible to track or approximate annual migrations out of a major.

The second choice of a data source that maximizes results while minimizing effort was statewide data repositories. This approach had been very successful in Virginia where the State Council of Higher Education for Virginia (SCHEV) supplied the data for all 23 programs in the pilot study for this project. SCHEV graciously agreed to extend their support by supplying updated statistics for the national study.

To enlist support and gain legitimacy, SHEEO, the nationwide State Higher Education Executive Officers association was asked to help persuade state higher education councils to provide the requested data. This approach yielded commitments from five states, three of which have complied thus far. Seven states responded with regrets that they did not have the data needed for this study. Two states offered to make their data available for a fee. In one case, the cost was estimated at a total of $1000 for the data from any number of colleges or universities in the state. The representative of one systemwide data repository declined because the request "results in a fair amount of data processing for us, which I am not inclined to take on." To date, approaching state councils of higher education has yielded a total of 30 data sets out of the 209 requested. An additional three states have promised to supply 19 data sets. With continued attempts to reach non-responding councils known to have student unit record data, the number of data sets may still increase.

Sponsored requests

The limited results yielded by appeals to statewide data repositories prompted another approach. We directly contacted the computer science chairpersons in the remaining institutions. Our rationale was that, as interested parties and members of the study institution, their sponsorship would carry more weight than an external request.

Email sent to department heads in the Spring of 2001 described the project, its endorsements, and the anticipated outcomes, and asked them to obtain the data. Based on prior experience with institutional researchers who denied requests because they did not fit with

previously produced reports, we suggested that an effective data source might be the information technology staff in the Registrar's office. In some cases, this suggested approach was successful. There were also departments that had administrators in possession of the necessary data, and in many other cases, departments found their Institutional Research Office to be the most appropriate source.

In conjunction with these sponsored requests, we created a website for the project. The website described the project and gave details about the data needed. It provided an example of the suggested file layout and defined terms for the project. Each email to computer science department heads included the address for this webpage as an easy reference (http://curry.edschool.virginia.edu/ITattrit/).

Perhaps because chairpersons were also asked to complete a questionnaire in the survey portion of this project, the sponsored request approach to gathering data for the dependent variable yielded only twelve commitments to provide data. Of these commitments, 5 data sets have arrived to date. In one case, the data have been delayed by requirements that a university's institutional review board consider the project. Despite the fact that the project's home institutional review board had already considered and approved the protocols, and that only aggregate data was being requested, a full review was conducted before the data could be sent.

Data-sharing consortia and Institutional Researchers

Twenty-three study institutions belong to a data-sharing consortium that focuses on retention of science, mathematics, engineering, and technology majors. The data member institutions supply

the consortium are similar, but not identical to, the data needed for this gendered attrition study. Thus, although data from the consortium would not meet the needs of this project, consortium members should have the capability to provide the data needed for this study.

Beginning in summer 2001 and continuing through winter, we directly contacted institutional researchers at non-responding colleges and universities. The email message contained a brief description of the projects' goals and endorsements, an outline of the requested data, and a link to the project website providing a more detailed description of the necessary data and background on the project. Members of the data-sharing consortium were alerted to the similarity of the requested data to the data they supplied the consortium.

The results from direct contact with institutional researchers were 27 data sets received and 11 promised but not yet sent. Of the responding institutional researchers, all but four were members of the data-sharing consortium. As was the case with state higher education councils and with chairpersons, most institutional researchers who were contacted did not reply. Follow-up contacts are continuing with non-respondents.

Among those institutional researchers who replied but declined to participate, by far the most common reason for not providing the data was a lack of sufficient resources. For example, one institutional researcher wrote,

> [W]e're drowning in deadlines and ever-expanding external demands, along with a massive system-wide changeover to [different software]--the effects of which I can describe in one word: Aaaarrggghhh! Our tiny staff has not been able to

> respond to requests like yours, which are perfectly reasonable and do represent areas of research that we've been trying to get to for our OWN campus's use!

Despite interest in the topic and desire to assist with this project, this office simply did not have the staff time.

The second most common reason given for not providing the data was that it had not been collected. For example, the following institution was only beginning to keep records of dropouts.

> I really doubt that we have anything of value. We are just starting to identify our dropouts generally and will probably be a year at that task . . .

Under these conditions, it would be a particularly time consuming task to locate annual lists of undergraduate computer science majors and graduating seniors, and match them from year to year. Simply not having the data in a useful format makes it hard to recognize it as suitable for this study, and more resource intensive to use if it is recognized. Despite the offer to have project staff work from annual lists of students by ID code, sex, level, etcetera, and convert them into the necessary format, no institution chose that option.

Results from collecting official data

After two years of continual attempts to gather official data for calculating gendered attrition rates, we achieved a compliance rate of 30%. The most productive source was the state higher education councils, which supplied 48% of the data sets received to date. Last call notices may result in some additional data before analyses are underway.

Regardless of the importance of these official data, gathering them has been the least successful aspect of this project. The disappointing results with this key element for analyzing departmental effects on the disproportionate loss of female students have delayed the project. Deadlines have been extended so that overworked institutional researchers can comply with our request, and so that new approaches for increasing participation can be employed. For example, several study institutions are part of a citywide university system. The system administration may serve as a source of data for all the non-responding institutions located in that city.

The challenges of the official data component of this project were identifying appropriate sources of the data, enlisting cooperation from database proprietors, tracking the status of requests, and working with variations in the data supplied by different sources. Governmental data sources were located with the assistance of the State Council of Higher Education for Virginia and SHEEO. Institutional sources were located by student assistants who searched webpages and called institutions to obtain names and email addresses. Enlisting cooperation was most effective when the proper data source was located; when the request was clear and details were provided, as accomplished by the sample layout on the project website; and when flexibility in deadlines and data content was offered. Tracking the status of requests required a separate database. This database contained an institution ID, name, position, email address and telephone number for each contact person, an identifier for the current contact, and notes describing the status of the request. Standardizing the data provided by different sources involved careful comparisons of definitions and calculations of proxy measures when the

requested information was not available. However these procedures improved our ability to obtain the official data, they were only marginally effective.

As with the other components of this project, the results from analyses of the official data offered some obvious benefits for participating institutions and for higher education. The benefits for institutions supplying the official data included experience with tracking the outflow of students from degree programs; a snapshot of conditions in one of their largest, most in-demand majors; and access to statistics measuring gender composition and migration patterns at other institutions. An example of the data products can be seen at the website presenting descriptive results from the statewide project at http://faculty.virginia.edu/attrition-cs-bio/, where statistics for computer science and biology departments in Virginia are shown. For higher education in general, the rewards included a contribution to understanding of the process by which men and women are segregated into particular disciplines, insight into the general relationship between departmental characteristics and undergraduate retention in particular programs, and a measure of gendered attrition rates that can be used to compare program outcomes across institutions.

Discussion

This case study described the methodology employed for a large multi-institutional research project. The goal of the project was to identify departmental factors that affect equality of male and female undergraduate retention in computing majors. Quantitative and qualitative data were collected from students, faculty, and chairpersons in the largest and/or most prestigious computer

science programs in the contiguous United States. In addition, official enrollment and disposition data were gathered for calculating measures of student migration from participating departments during a six-year period.

The project methodology presented different degrees of challenge, and had varying levels of success as shown in Figure 2. Data collection by interview and survey was demanding but profitable. In comparison, collection of data from official sources was more difficult and met with less success.

The data collection methods for the qualitative portion of this project followed established procedures, and were productive. Data were collected directly from individual participants at study departments through interviews and focus groups. The departmental participation rate for this component of the project was high (82%).

Most chairpersons were open to hosting site visits, graciously making physical arrangements and encouraging members of their departments to participate. Although the level of individual participation varied across departments, the overall number of department members interviewed was high (325). As a result, the interviews and focus groups covered a large number of individuals at a variety of institutions. The data they generated made a valuable contribution to the quality of the subsequent survey, resulted in reports that departments could use for self-evaluation and comparison with other computer science programs, and produced a wealth of observations that could stimulate new hypotheses and illustrate statistical results. Because the data were obtained from many institutions of various types and locations, they are likely to

represent the range of conditions present in large academic computer science departments and depict the features common to most of these departments.

The data collection methods for the survey portion of this project also followed established procedures, and they were also very productive. Data were collected directly from individual participants at study departments through a questionnaire that employed the Internet, mail, or telephone. The response rate for this component of the project was high (67%).

The survey data measured the distribution of characteristics and practices in computer science by department, by sex, and by a variety of other institutional, program, and individual characteristics. The example of factors affecting faculty turnover showed how multi-institutional research can offer significant information relevant to institutional policy. Because many faculty supplied the data for this result, it can be legitimately claimed that this finding represents the situation for the average faculty member in a large computer science department within a two percent margin of error. Thus, the finding applies broadly and cannot be dismissed lightly as unique to a particular institution. Furthermore, when analyses by department are conducted, it will be possible to discover which departments deviate from the norm and to investigate the reasons for those deviations. Such comparisons across departments will offer abundant information for self-assessment and for deepening understanding of factors relevant to collegiate preparation for computing careers.

The rates for the survey's three response modes demonstrate that this study's population was quite appropriate for an online questionnaire. A large majority of participants responded online (90% of those responding used this method), resulting in a tremendous cost saving for the

project. The boon of reduced cost, together with easy communication and access to public information illustrate how technological advances have made multi-institution research more feasible. However, one should not interpret these survey methodology results as having implications for response rates in general. The three response modes for this survey were not offered simultaneously, so comparing response rates is not legitimate. The web survey was available first and for the entire study period, but mail and telephone were only offered after responses via the web declined. Thus while the Internet provided many benefits, raising response rates was not necessarily one of them.

The data collection methods for the official data portion of this project had no well-established procedures to follow, and were the least productive of the project methodologies. Official data for calculating departments' gender composition and gendered attrition rates were collected from government sources or directly from study institutions. This process had the lowest response rate of the project (30% participation rate).

Following a similar process to that used for implementing a survey, potential data sources were located and contacted. Request status was tracked and non-response was followed with additional contact attempts as well as contact with alternative sources. Among those who responded but declined to participate, the chief reason was reluctance to commit scarce resources for extracting data. Among those who did participate, statewide data repositories were not the most cooperative, but they were the most productive (48% of the data sets came from this source), making them a highly efficient data source for multi-institution researchers. The next

most productive, and the most cooperative source was individual institutional research offices (44% of the data sets provided).

The current low response rate for official data jeopardizes the project's ability to achieve its primary goal of measuring the links between departmental factors and their gendered attrition outcomes. If only 30 percent of the study departments are represented, there is serious potential for bias. The gendered attrition rates of participating departments might not accurately represent rates for the typical large computer science department. Further analyses will determine how representative the study data actually are, but increasing the number of participants in this aspect of the project would reduce the margin of error for the statistical findings based on the official data.

This case study illustrated several important issues regarding the current condition of multi-institution research. Examining the process and outcomes of this project demonstrated the benefits, challenges, and persisting impediments to research involving many colleges and universities.

The contributions of this project confirmed the worth of multi-institution research already demonstrated by other research of this type. (For examples of other research, see Astin and Astin, 1992; and Strenta, 1994.) In the case of this study, findings can help to meet the demand for computer professionals, promote female participation in a rewarding career field, and deepen our understanding of gender segregation processes in higher education. The value of these findings is enhanced when they can be generalized to all large undergraduate computing programs, a feature that requires inclusion of multiple and diverse institutions.

This case study also substantiated methodological advances in data collection from multiple institutions. Academic computer science is particularly suited to study via cost-saving web surveys. The use of email facilitated communication between both researchers and potential participants. Web access to academic department sites, contact information for institutional research staff, and government data sets made it possible to gather certain data without response burden. In several cases, the existence and cooperation of statewide sources drastically minimized the effort to acquire data.

Despite these benefits and advances, this case study also demonstrated that data availability still impedes the conduct of multi-institution studies. Restricted access, cumbersome data storage, and inadequate resources have yet to be overcome.

Finally, this case study suggested that the most efficient means of disseminating data for multi-institution research is through government sources. These sources can provide access to high quality data from large numbers of institutions, minimize the data collection efforts of higher education researchers, and relieve institutional researchers of the burden imposed by extracting data for idiosyncratic requests.

A national commitment to quality education requires a companion commitment to collect, maintain, and disseminate the data that facilitate research and inform decision-making in higher education. Government support can provide the necessary resources. Without this support, multi-institution research will fail to thrive, and we will forfeit the educational and social progress it generates.

References

Astin, Alexander W. and Helen S. Astin. 1992. Undergraduate Science Education: The Impact of Different College Environments on the Educational Pipeline in the Sciences. Higher Education Research Institute. Graduate School of Education. University of California, Los Angeles.

Borden, Victor M. H.. 1995. Harnessing New Technologies for Student Tracking. *New Directions for Institutional Research*. Number 87.

Creswell, John W., Jay L. Chronister, and Martha L. Brown. 1991. The Characteristics and Utility of National Faculty Surveys." *New Directions for Institutional Research*. Number 69.

Dillman, Don A. 2000. *Mail and Internet Surveys: the tailored design method*. New York, J. Wiley.

Ewell, Peter T. Editor's Notes. 1995. *New Directions for Institutional Research*. Number 87.

Firnberg, James W. 1991.Access and Use of Federal Data Through NSF's CASPAR System. *New Directions for Institutional Research*. Number 69.

Hackett, E. Raymond. 1996. Creating a Cost-Effective Data Exchange. *New Directions for Institutional Research*. Number 89.

Lenth, Charles S. Editor's Notes. 1991. *New Directions for Institutional Research*. Number 69.

Pavel D. Michael and Mark Reiser, 1991 Using National Data Bases to Examine Minority Student Success in Higher Education. *New Directions for Institutional Research.*. Number 69.

Russell, Alene Bycer. 1999. The Status of Statewide Student Transition Data Systems: A Survey of SHEEO Agencies. Report issued by the State Higher Education Executive Officers Association. May 1999.

Russell, Alene Bycer, and Mark P. Chisholm. 1995. Tracking in Multi-Institutional Contexts. *New Directions for Institutional Research*. Number 87.

Sapp, Mary M. 1996. Benefits and Potential Problems Associated with Effective Data-Sharing Consortia. *New Directions for Institutional Research*. Number 89.

Strenta, A. Christopher, Rogers Elliott, Russell Adair, Michael Matier, and Jannah Scott. 1994. Choosing and Leaving Science in Highly Selective Institutions. *Research in Higher Education* Vol. 35, No. 5

Trainer, James F. 1996. Coming Aboard: Making the Decision to Join a Data-Sharing Organization. *New Directions for Institutional Research*. Number 89.

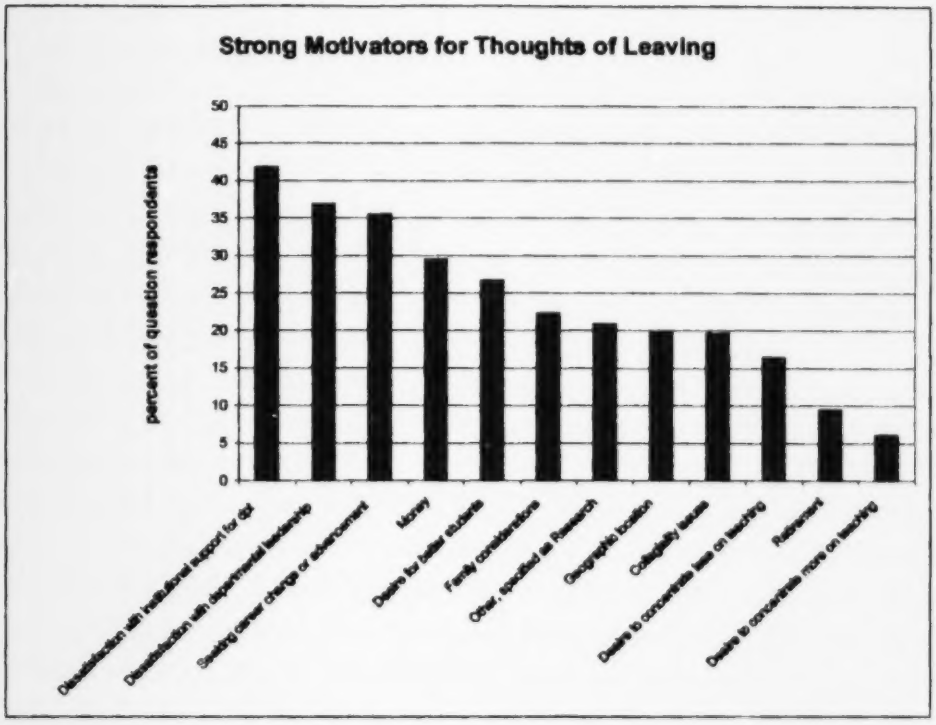
Vinsonhaler, Jeane C. and John F. Vinsonhaler. 1991. Developing a Data Base for Monitoring and Improving Quality. *New Directions for Institutional Research*.. Number 69.

Young, Denise York and Lawrence J. Redlinger. 2001. Modeling Student Flows through The University's Pipelines. Presented at 41st Forum of the Association for Institutional Research. Long Beach, California. June 5, 2001.

Author Notes

This material is based upon work supported by the National Science Foundation under grant number EIA0089959. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Figure 1. Sample Result from Survey Data.



BEST COPY AVAILABLE

Figure 2. Summary of Response to Study Methodologies

Study Response Rates by Methodology

Data Type	Response Rate	Sample n
interview/focus group	82%* (325 participants)	22 departments
survey	67%	2735
web	90%	
mail	8%	
telephone	2%	
official enrollment and disposition statistics	30%	209
government	48%	
cs department	8%	
institutional research	44%	

*This is the rate at which departments agreed to host interviews. Internally, departments varied in their success with recruiting participants.

BEST COPY AVAILABLE

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

☑ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☐ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").